Generation of Folk Song Melodies using Bayes Transforms

Chris Thornton
COGS/Informatics
University of Sussex
Brighton
BN1 9QH
UK
c.thornton@sussex.ac.uk

March 9, 2011

Abstract

The paper introduces the 'Bayes transform', a mathematical procedure for putting data into a hierarchical representation. Applicable to any type of data, the procedure yields interesting results when applied to sequences. In this case, the representation obtained implicitly models the repetition hierarchy of the source. There are then natural applications to music. Derivation of Bayes transforms can be the means of determining the repetition hierarchy of note sequences (melodies) in an empirical and domain-general way. The paper investigates application of this approach to Folk Song, examining the results that can be obtained by treating such transforms as generative models.

1 Introduction

While music is often analyzed in terms of specifically musical concepts (e.g., Schenker, 1935/1979), there is a long tradition of interpretation using generic concepts. Lerdahl and Jackendoff note that 'Medieval theorists justified their [musical] constructs partly on theological grounds' (Lerdahl and Jackendoff, 1983, p. 1), while also observing the extent to which physical and philosophical principles have been used. Modern work more commonly cites mathematical or statistical principles. Indeed, in Lerdalh and Jackendoff's view, the search for a 'mathematical foundation for the constructs and relationships of music theory' (ibid., p. 1) is a fundamental issue for research.

An important stimulus for generic analysis of music is modern information theory (Shannon, 1948; Shannon and Weaver, 1949). In contributing the concept of informational *entropy*, this framework provides the means of applying

a mathematical measure of structure to musical data. In an early proposal, entropy is seen as having a direct connection with musical meaning. Meyer conjectures that 'the psycho-stylistic conditions which give rise to musical meaning, whether affective or intellectual, are the same as those which communicate information' (Meyer, 1957/1967, p. 412). With the relationship envisaged to reflect an underlying 'law of entropy', connections are then posited between informational properties and various aspects of musical content.

Other researchers deploy the concept in more instrumental ways, e.g., for modeling generative composition of nursery tunes (Pinkerton, 1956), classifying musical style (Youngblood, 1958; Knopoff and Hutchinson, 1981; Knopoff and Hutchinson, 1983) and for contrasting differently styled passages in a single piece (Hiller and Fuller, 1967). Such approaches face the challenge, however, of obtaining reliable estimates for the probabilities on which entropy measurement is based.

In principle, the procedure is entirely objective: frequencies obtained by sampling relevant data are converted to probabilities, producing a well-defined measurement of entropy. Intuitively, entropy is the degree to which a set of probabilities are evenly distributed. Formally, the entropy of distribution P is

$$-\sum_i P_i \log P_i$$

where P_i is the probability of the i'th outcome.

A problem arises, however, if salient probabilities cannot be obtained by sampling frequencies (Temperley, 2007). As Meyer notes 'Not all probabilities embodied in a musical composition are determined by frequency' (Meyer, 1957/1967, p. 422). Estimation through frequency sampling in such situations is necessarily ruled out. Probabilities may also be intrinsically subjective in nature. Again, this generally means they cannot be obtained by sampling (Cohen, 1962). In such cases, there remains the possibility of informal evaluation. As in Shannon's study of English (Shannon, 1951), probabilities may be estimated by polling human judgement. But the entropy measurement obtained is then only as reliable as the estimates on which it is based.

There are considerable challenges, then, confronting the attempt to apply information measures to music, and these relate particularly to determination of probabilities (Margulis, 2008). The expectation that entropy might provide a generic measure of musical structure has also diminished somewhat, with awareness increasing of the significance of hierarchy and grouping. Whereas entropy quantifies distributional uniformity, the structural factor of particular significance in music is often held to be *repetition* (Schenker, 1954; Lerdahl and Jackendoff, 1983; Bent and Drabkin, 1987). Indeed, some researchers see repetition as fundamental. In Dannenberg and Hu's view, for example, 'musical structure is signaled by repetition' (Dannenberg and Hu, 2003, p. 153). But repetition is something the distribution-oriented entropy measure cannot directly reflect, particularly given the way patterns of repetition can exist in myriad different forms at multiple levels of description (Meredith *et al.* 2002). (A more detailed discussion of factors relating to modeling of repetition is provided in Section 4.)

On the face of it, modeling patterns of repetition and grouping necessitates some form of hierarchical representation, such as Schenkerian analysis (Schenker, 1935/1979; Schenker, 1954), the grammatical approach of Lerdahl and Jackendoff (1983), or the preference rules of Temperley (2001). But application of information theory is not necessarily ruled out. Use of entropy measurement can be combined with hierarchical forms of learning. Most straightforwardly, this involves utilization of features for purposes of deriving abstractions over musical data. Entropy's capacity to measure relative uncertainty (rather than distributional uniformity) can then be utilized for purposes of finding, refining, or selecting that combination of features that represents relevant data with least uncertainty.

A prominent example of this strategy is the multiple-viewpoints framework (Conklin and Witten, 1995; Conklin, 2002; Conklin, 2002; Conklin and Anagnostopoulou, 2006; Bergeron and Conklin, 2007). (Coverage in this section focuses on entropy-based methods as they relate to the Bayes transform introduced in Section 2. Section 4 deals more fully with how these methods relate to the results obtained.) This approach involves 'deriving individual expert models for any given representational viewpoint and then combining the results obtained from each model' (Pearce *et al.* 2005, p. 295). Other approaches use variants of the entropy measure to similar ends. In Temperley's approach (Temperley, 2007; Temperley, 2008), Bayesian models conditioned on pre-specified features are selected on the basis of *cross-entropy*. Such measurements reflect the relative probability of data given a specific model. Conklin's use of 'segment classes' (Conklin, 2006) is along similar lines. In Bod's parse-based approach to phrase prediction (Bod, 2001), application of the 'predictability' principle plays a comparable role.

Another way of utilizing entropy measurements towards representation of structure is explored in the IDyOM model of (Pearce and Müllensiefen, 2008; Pearce *et al.* 2008). This approach focuses on the relative uncertainty of future note events given knowledge of past events. By observing the way conditional uncertainty varies over a musical sequence, it becomes possible to predict phrasing boundaries. Based on use of *N*-grams, this approach faces the general problem (revisited below) of knowing what length (order) of *N*-gram will produce best results. As Pearce and Müllensiefen note, 'Low-order models fail to provide an adequate account of the structural influence of the context. However, increasing the order can prevent the model from capturing much of the statistical regularity present in the training set' (Pearce and Müllensiefen, 2008, p. 91). The solution in IDyOM is to learn a system of weights, and then use them to combine differently sourced probabilities.

A broad range of methods are available, then, for deployment of entropy-measurement in representation-construction. Often, those facets which particularly promote representation of structure (e.g., the 'derived types' of the multiple-viewpoints approach) have to be pre-specified, out of domain knowledge, however. The result is then a model in which only some aspects of the representation are the result of entropy minimization, and its impact on the modeling cannot be easily determined. More practically, there is the difficulty

of arriving at the requisite pre-specified entities, i.e., the problem of determining what domain knowledge will be of value. Ideally, such issues should be resolved through, rather than separately to, any optimizations being performed.

Of potential interest in this context is an information-theoretic operation the present paper dubs the 'Bayes transform'. Like the methods noted above, this deploys entropy measurement towards formation of structural representations. But rather than the higher levels of representation being built-in, in the form of user-specified features, they emerge through recursive application of the operation itself. The higher levels of representation are built-up, step by step, by application of the operation to results it has previously produced.

Essentially a way of putting a data sequence into an informationally efficient, hierarchical representation, the operation can also be seen as a way of deriving a series of diminishingly complex Bayesian models, in which each model references a distinct level of organization in the data. More straightforwardly, it is a way of building a generalization hierarchy. What makes it relevant to the case of music is its ability to capture patterns of commonality. It is a mathematical consequence of the formulation that the hierarchy obtained represents patterns of duplication at successive levels of generalization. This becomes of interest where the process is applied to musical data, such as a melodic sequence. In this case patterns of duplication may constitute patterns of repetition. The general effect obtained is then the capturing of patterns of repetition at multiple levels of hierarchical organization.

To the extent that musical qualities are constituted in such structured regularities, deriving the Bayes transform becomes of potential use as a data-driven modeling strategy. Should qualities of metrical structure be so constituted, derivation of transforms may be of use in representing rhythm. Should qualities of phrase structure be so constituted, the process is potentially of use in predicting boundaries. The mechanism may then have potential as a context-free way of modeling certain forms of musical phenomena.

Does it have any value in practice, however? The research described herein aims to assess the situation through experiments in the domain of Folk Song, with data drawn from the Essen corpus (Schaffrath, 1995). This database provides digital encodings for a large number of melodies in various styles, taken from different regions and countries of Europe. Bayes transforms of melodies taken from this corpus will be examined for their potential to predict phrase boundaries. More particularly, attention will be given to the degree to which such representations can be used generatively, i.e., as structures from which variants of particular melodies or styles can be generated.

However, the basis on which this is done is somewhat unusual. The Bayes transform is *not* set out as a method for modeling music. As an information-theoretic method for determining repetition structure, it cannot fulfil this role. The main aim of the paper is to examine the extent to which hierarchical modeling of repetition can provide a vehicle for generative applications. There is thus no comparison of Bayes transforms with analyses derived from existing areas of music theory. There is no attempt to determine how well Bayes transforms capture musical nuances, features or properties. This is not a proposal for a new

approach to music analysis. Rather, it is an examination of areas of overlap between this task and a certain form of statistical analysis.

The paper has five main sections. The following section sets out the mathematics of the Bayes transform, illustrating its application to text sequences, and examining some of the ways of viewing the hierarchical structures that are obtained. Section 3 explores use of transforms for modeling of melody, particularly with regard to generative modeling of Folk Song melodies. Section 4 discusses related work and Section 5 presents a summary and some conclusions.

2 Formalization

The operation here termed the 'Bayes transform' provides the means of putting data into an informationally optimal, hierarchical representation (cf. Thornton, 2010). Applicable to any dataset regardless of constitution, the method is particularly of use where there is a need to exploit patterns of commonality at multiple levels of organization. We start by envisaging data D to be a set of symbols drawn from an alphabet of n elements. Denoting the number of symbols in D as $|D|$, the total information content of the data is

$$I(D) = |D| \cdot \log n \tag{1}$$

No constraints are placed on the structure of D. As a simple illustration, D might be the sequence 'X Y X Y Z'. If symbols are drawn from an alphabet of 26 elements, the total information content is then

$$|D| \cdot \log 26 = 5 \cdot 4.7 = 23.5 \text{ bits}$$

Logs are taken to base 2 here and throughout.

No constraints are placed on the constructs that D exhibits. Where two or more constructs share the same structure, their union can be referenced in certain ways, however. Specifically, if x represents a particular union of constructs, $|x|$ is the number of symbols in the (common) structure, and x_i is the set representing the choice of symbols for the i'th element of the structure.

Continuing to view the data as the sequence 'X Y X Y Z', constructs might be taken to be subsequences, such as 'X Y'. Among the three-element subsequences, we would then have 'X Y X' and 'X Y Z'. If x represents the union of these two, the relevant choices would then be $x_1 = \{\text{X}\}$, $x_2 = \{\text{Y}\}$ and $x_3 = \{\text{X, Z}\}$, where subscripts correspond to indexes in the obvious way. The shorthand used to represent this union would be 'X Y X/Z'

Building on these definitions, it is possible to introduce D', denoting a *reconstruction* of D. This is a modification of D in which some of its constructs are replaced with symbols representing unions. Replacement is deemed to be possible just in case the construct is within the represented union.

In the case of the sequence 'X Y X Y Z', we might have the reconstruction '$0 X Y Z', where $0 is a symbolic label for the union 'X Y/Z'. This is possible since the two element construct 'X Y' is within the union 'X Y/Z' (which combines 'X

Y' and 'X Z'). Where replacements introduce choice, there is loss of information, i.e., increase in uncertainty. This can be precisely quantified. The information loss (i.e., uncertainty) resulting from a replacement involving a particular union x may be calculated as

$$H(x) = \sum_i \log |x_i| \tag{2}$$

Equivalently, it may be calculated as the log of the combinatorial product of x's choices:

$$H(x) = \log \prod_i |x_i| \tag{3}$$

The total information lost in a reconstruction can then be calculated by summing the information losses associated with its symbols:

$$H(D') = \sum_i H(D'_i) \tag{4}$$

Here, $H(D'_i)$ is zero if D'_i is an original symbol, and the information loss of the represented union otherwise.

The total symbol cost of a reconstruction (the total number of symbols used) can also be defined. It is the number of symbols used in the modification itself, added to the total number of symbols used in referenced constructs. This cost is denoted $c(D')$:

$$c(D') = |D'| + \sum_{x \in D'} |x| \tag{5}$$

In this formula, $x \in D'$ enumerates the set of constructs referenced by D'. As an illustration, the symbol cost of the reconstruction '\$0 X Y Z' is

$$4 + 2 = 6$$

Combining the reconstruction loss with the reconstruction cost, it is then possible to define the informational efficiency of a reconstruction, i.e., the mean information content of symbols. This is the information content of the original data less the reconstruction's total loss, divided by symbol usage:

$$\bar{I}(D') = \frac{I(D) - H(D')}{c(D')} \tag{6}$$

The informationally optimal reconstruction of D is then that reconstruction that maximizes mean information. This is denoted $r(D)$:

$$r(D) = \operatorname{argmax} \bar{I}(D') \tag{7}$$

Note that the mean information of $r(D)$ can be no less than that of D itself. Were this to be the case, D would be its own optimal reconstruction. Given $r(D) \neq D$, it must also be the case that

$$\bar{I}(r(D)) > \bar{I}(D) \tag{8}$$

which further implies that

$$c(r(D)) < |D|.$$

Increasing the mean content of symbols above the level they have in D itself must involve reducing their number. The optimal reconstruction must use a lesser number of symbols than D itself.

Maximizing mean symbol information, the reconstruction $r(D)$ is the most informationally efficient representation of D's content that can be derived in terms of D's native constructs. Its interesting property from the modeling point of view is the potential inclusion of generalizations. Note these are not the products of an explicit 'generalization procedure'. Their introduction serves the goal of maximizing information, and is progressed only up to the point at which informational costs (in terms of increased uncertainty) are balanced by the informational benefits of reduced symbol usage. Should it be possible to achieve such benefits without paying *any* price in terms of lost information (perhaps because the dataset presents explicit duplication), no generalization may be produced.

As an illustration of how optimal reconstruction works in practice, consider the sequence 'a b c a d c a e c'. On the assumption that constructs of this dataset are subsequences, the optimal reconstruction of the dataset turns out to be '\$0 \$0 \$0', where \$0 represents the generalization 'a b/d/e c'. The choices represented in this correspond to alternations in the obvious way: the choice for the first symbol is {a}; the choice for the second symbol is {b, d, e} and the choice for the third is {c}. Given $\log 1 = 0$, all the terms relating to singleton-set choices drop out: the total uncertainty of the reconstruction is simply

$$\log 3 \cdot 3 = 4.75 \text{ bits}$$

Assuming original symbols are drawn from an eight-symbol alphabet, the information content of the original dataset is $9 \cdot \log 8 = 27$ bits. With the symbol cost of the reconstruction being $3 + 3 = 6$ (Equation 5), mean information is found to be

$$\frac{27 - 4.75}{6} = 3.71 \text{ bits}$$

In this case, derivation of the optimal reconstruction serves to increase mean information by 0.71 bits, with the number of symbols being reduced by 1/3. These effects are facilitated by the existence of approximate repetitions in the source dataset. These enable the dataset to be broken into generalized subsequences at a modest loss of information. Referencing the generalization in question, the original dataset can then be represented as a three-element sequence, with a total information loss of 4.75 bits.

Focusing on the generalization aspect of the process, we might plausibly treat '$0 $0 $0' as a kind of inductive hypothesis, on which basis use of information theory in the derivation might be seen as a learning heuristic. Another possibility is to see the process as a form of lossy data compression (Held, 1987), involving the conversion of a relatively redundant encoding into a more compact form. A third possibility is to see the process as a mechanism for finding an optimal tradeoff between the costs of generalization and the benefits of parsimonious encoding.

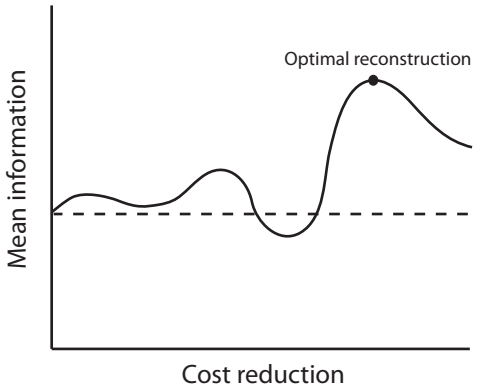

Figure 1: The information/reconstruction tradeoff.

Offering a visual illustration of this third interpretation is the graph of Figure 1. The horizontal axis, here, represents the achieved reduction in symbol cost. The vertical axis represents the mean information achieved. The dashed line is the informational baseline, i.e, the information content of original symbols. The curve itself then represents the maximum mean information that can be achieved (through reconstruction) for each level of information loss (i.e., generalization). The optimal reconstruction is then the highest point — or, in general, points — on the curve.

2.1 Recursive refinement

Finding the informationally optimal reconstruction of a dataset can serve the goals of generalization and/or data compression. But the potential of the process is more fully realized when it is deployed in a *recursive* manner. The end-product of a single reconstruction is a structure of symbols. This is another dataset, for which we can derive a second, optimal reconstruction. The effect, in this case, is to exploit generalizations over symbols representing generalizations at a lower level of organization. Taking the process forward recursively, we then obtain a *series* of reconstructions, the constructs of which capture generalizations at increasingly coarse-grained levels of organization.

Letting optimal reconstructions now be called 'refinements', the optimal reconstruction of an original dataset can be termed the 'first refinement', the optimal reconstruction of the reconstruction the 'second refinement', and so on. Any dataset then has a first refinement, second refinement, third refinement etc., with the total number depending on the constitution of the data. The requirement for an optimal reconstruction to show an increase of mean information (Equation 8) imposes a limit, however. The entire hierarchy of optimal refinements can then be defined using the following, recursive formula:

$$D^n = r(D^{n-1}) : \bar{I}(D^n) > \bar{I}(D^{n-1}) \tag{9}$$

Labeling the original dataset D^0, this formula specifies the constitution of the first refinement D^1, the second refinement D^2, third refinement D^3, and so on, up to the n'th refinement D^n. The value of n is the *last* level at which the optimal reconstruction obtained has mean information greater than that of its source dataset. Beyond this level, further refinement is ruled out. The representation obtained at this level is thus the *root refinement* of the dataset.

Since optimal reconstructions are not necessarily unique, there may be more than one refinement hierarchy, however, and thus more than one root refinement for any given dataset. There is also the possibility for determining an earlier cutoff in terms of a minimum requirement on information increase, or a minimum requirement on cost reduction. Meeting this requirement would then produce a subjectively defined root refinement.

2.2 Illustration

Illustrating a complete derivation of possible refinement levels is the schematic of Figure 2. This presents refinements obtained for a dataset named 'eg1', which is the sequence of 12 characters 'b c d b e d b f d b g d'. (Spaces are used here and throughout as separators.) Possible primitives are taken to be the eight characters {a, b, c, d, e, f, g, h}, with constructs treated as subsequences in the usual way.

The four levels of the tree-structure of Figure 2 correspond to the four levels of refinement obtained, with higher-level refinements appearing higher in the figure. At the lowest level, we see the dataset itself, represented as a sequence of 12, oval-shaped nodes. Each one of these nodes encloses a representation for a particular element from the dataset sequence. In general, ovals represent symbols, however, with their contents representing the referenced construct. Arcs, where shown, point to the locations of relevant constituents. With eight primitive symbols in use, each element of the original dataset has an information value of $\log 8 = 3$ bits. Comprising 12 symbols in all, the dataset then has a total content of 36.0 bits. (Annotations to this effect appear on the right of the figure.)

At the first level of the hierarchy, we see the initial refinement of the data. As in the previous example, this exploits the presence of approximately repeated constructs. In this case, the repeats commence at indexes 0, 3, 6 and 9, while

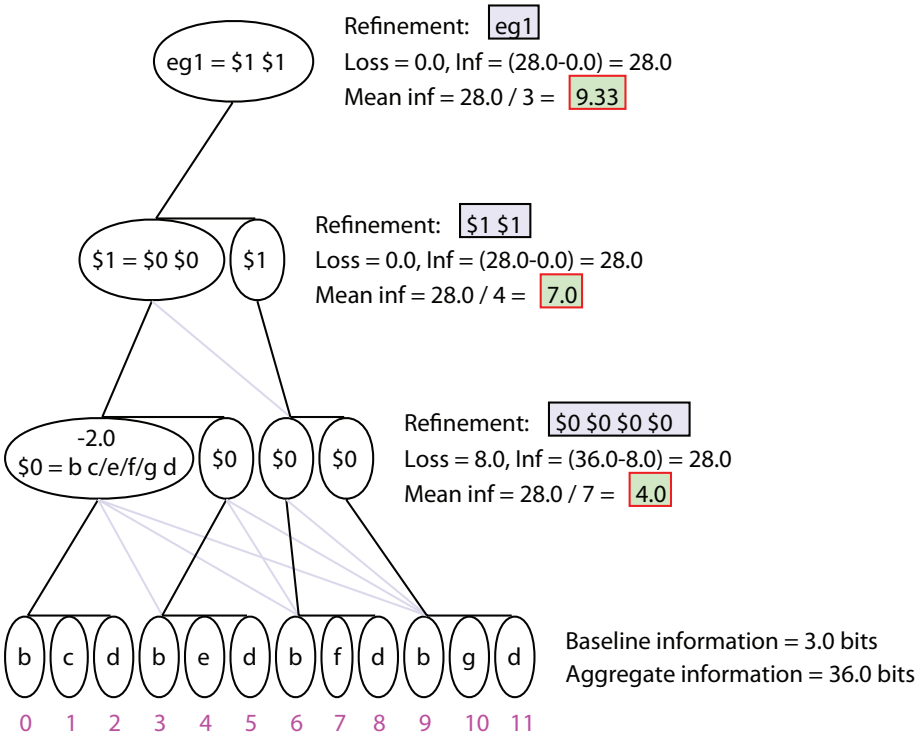

Figure 2: Refinement hierarchy for the sequence 'b c d b e d b f d b g d'.

the generalization in question is 'b c/e/f/g d'. With the original dataset recoded as four repetitions of this construct (here labeled '$0') an information loss of 8.0 bits is incurred, arising as 2.0 bits per use of the construct. Subtracting this aggregate loss from the original 36.0 bits, we have a residual content of 28.0 bits. Dividing by the seven symbols in use (comprised of the three symbols of the construct itself, and the four uses of $0), we obtain an information per symbol of 4.0 bits. Precisely one bit above the baseline of 3.0 bits, this is the maximum mean information that can be achieved at this level.

Reconstruction involves the introduction of specific degrees of choice in specific symbols. Determining the optimal reconstruction in a given case is thus accomplished by searching through possible combinations of choice to find the one that yields highest mean information. Where the number of data make this prohibitively costly, it is necessary to resort some form of hill-climbing search, in which case, the reconstruction is not guaranteed to be optimal. (Most but not all of the examples below were derived using non-heuristic search.)

At the next highest level of the hierarchy, the refinement obtained embodies no information loss at all. This exemplifies the point made previously, about generalization being a side-effect. The entire 28.0 bits of content from the first

refinement is retained. Repetition of the subsequence '$0 $0' is exploited, however, through introduction of the construct labeled $1. This yields a further increase in mean information. Finally, at the highest level, the root of the hierarchy is obtained. At 9.33 bits per symbol, mean information at this level is 6.33 bits above the baseline.

All three refinements in this hierarchy maximize mean information content. Only the second refinement, however, exhibits any effects of generalization. In fact, $0 is the only generalized construct in the transform. In the case of more complex datasets, refinement hierarchies typically exhibit multiple generalizations at several levels of organization. These are of more obvious application to modeling.

2.3 Characterizing the refinement process

Comments have already been made about ways to interpret the results of optimal reconstruction. One view focuses on the classificatory aspect. Recognizing that optimal reconstruction may involve generalization, there is the potential to view the mechanism as a kind of inductive learning procedure. This interpretation carries over to the recursive case, although the objective must then be seen as production of a complete hierarchy, rather than a single level of generalization. The assessment is not completely accurate, however, since the inductive element in the process is essentially a side-effect. Depending on the data, the refinement hierarchy may present no effects of generalization at all.

Another interpretation focuses on the reductive aspect of the process. Recognizing that optimal reconstruction has the effect of producing more compact representations, there is the potential to view the process as a form of data compression. Where generalized constructs are referenced, the compression can be viewed as 'lossy'. Otherwise it is 'lossless'. Applying this interpretation to the hierarchical case is less natural, however. The result of the process is a series of reconstructions, whose total cost may substantially outweigh that of the original dataset. Recursive refinement cannot be regarded as a data compression method in itself, then. At best it can be viewed as a way of iterating a compression function for purposes of producing a series of diminishingly costly encodings.

Combining the compression and induction interpretations within a single view is also a possibility. A long tradition of work has emphasized a connection between simplification and induction (Wertheimer, 1923/1938; Solomonoff, 1964b; Solomonoff, 1964a; Chater, 1999). The idea is central to the paradigm of Minimum Description Length (MDL) learning (Rissanen, 1978; Rissanen, 1987), while being well known as the principle of Occam's Razor (Li and Vitányi, 1997, p. 317; Blumer *et al.* 1987). But attempting to interpret recursive refinement as a way of promoting a fundamental 'Occam principle' is obstructed by the effect of the process, which may be to increase rather than decrease the number of data in play.

A preferable approach involves use of Bayesian concepts. Here, we focus on the way information refinement benefits representation of conditional probabil-

ity. Since refinement maximizes the information content of its symbols, it can also be seen as *minimizing* expected uncertainty about the source dataset, when that dataset is viewed as the conditional product of the reconstruction. Specifically, the optimal reconstruction implicitly minimizes the expected uncertainty

$$-\sum_i P(D^n_{\langle i\rangle}) \sum_j P(D^{n-1}_{\langle j,i\rangle} | D^n_{\langle i\rangle}) \log P(D^{n-1}_{\langle j,i\rangle} | D^n_{\langle i\rangle})$$

Here, $D^n_{\langle i\rangle}$ is the i'th reference of the reconstruction and $D^{n-1}_{\langle j,i\rangle}$ is the j'th constituent of the construct represented by $D^n_{\langle i\rangle}$. Previously, the former was taken to be a symbol representing a construct made up of symbols from the source dataset. But we can also look at it the other way around, seeing the symbols in the source dataset as the conditional implications of the construct. On this basis, maximizing the informational efficiency of the reconstruction is equivalent to forming a set of Bayesian conditions which will maximize the conditional probability of the source data.

Viewing recursive refinement in this Bayesian way, the hierarchy is revealed to be a *spectrum* of optimal representations. Each refinement captures generalizations at a specific level of hierarchical organization, while also maximizing the conditional probability of its source data. (This is particularly inspired by Chaitin's proposal that information theory can be used to produce a 'kind of "spectrum" or "Fourier transform" ' (Chaitin, 1979, p. 88).) Derivation of the refinement hierarchy is then naturally seen as a way of converting data into a hierarchical form that is rich in conditional relationships, both with regard to itself and with regard to levels of organization in the originating dataset. Seemingly the most veridical among the possible interpretations, this will be the one used below. The term 'Bayes transform' is then introduced as a convenient nmemonic.

3 Transforms of musical data

Attention now turns to derivation of Bayes transforms from data representing melodies. Examples will involve transforms of note sequences, and assessment of these representations for generative applications. In all cases, derived datasets will be sequences of note codes, and constructs will be considered to be subsequences of up to four elements.

The first example to be examined involves the opening theme of Mozart's G Minor symphony K.500 (Figure 3). A frequently considered case in the literature (e.g., Lerdahl and Jackendoff, 1983, p. 37; Bod, 2001), this is of particular interest due to the prominent use of repetition in the opening bars. In fact, the initial nine notes break down into three, identical repeats. Taking the opening theme of the piece to comprise the pitch-class sequence 'Eb D D Eb D D Eb D D Bb Bb A G G F Eb Eb D C C', the Bayes transform derived from this data is shown in Figure 4.

The transform is a structure of three refinement levels, with generalizations captured in symbols \$0 and \$4. Some musically interpretable features are visi-

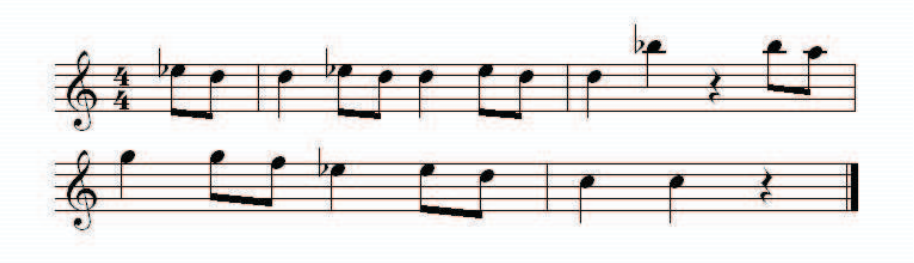

Figure 3: Initial melody of Mozart's G Minor Symphony, K. 550.

ble. The explicit repetition of the 'Eb D D' motive at the outset of the sequence is generalized by symbol $0, while $4 serves to (implicitly) group its three repeats. This sub-structure reflects modeling of the initial phrasing of the piece. Indeed, it conforms to the structure that Lerdahl and Jackendoff suggest is normally heard for this part of the theme (Lerdahl and Jackendoff, 1983, p. 37). However, construct $0 also generalizes the later fragment 'Eb D D', cutting across any sensible interpretation of phrasing. The generalization referenced by $4 also implicitly groups a phrase made up of the initial three repeats, with an uninterpretable entity comprising a mixture of subsequences. Derivation of the Bayes transform in this case illustrates the way in which exploitation of patterns of repetition may or may *not* lead to musically interpretable representations.

3.1 Modeling Folk Song phrasing

A useful testbed for evaluating prediction of phrasing is provided by the Essen Folk Song corpus (Schaffrath, 1995). This offers digital representations for the melodies of more than six thousand songs taken from different parts of Europe. (No particular position is taken on the degree to which the subsets of the corpus represent ethnographic groupings, or on the integrity of the phrasings that the corpus suggests.) In the version of the corpus currently supported at kern.ccarh.org, data are encoded in a format that represents pitch and rhythm values, time signatures and other information (Huron, 1999). The data also include phrase boundaries for the melodies, and it is this aspect that has facilitated work on prediction of phrasing (e.g., Hillewaere *et al.* 2009; Pearce *et al.* 2008; Bod, 2001; Temperley, 2001; Temperley, 2007).

To evaluate the degree to which Bayes transforms express phrasing analyses of Folk Song melodies, a study was carried on data taken from the corpus. The aim was to derive Bayes transforms for specific melodies, and to assess the degree to which the suggested phrase boundaries are also observed in transforms. Some pre-processing was necessitated in order to derive data in the form of pure sequences. Melodies in the corpus are represented in **kern format, which is a symbolic representation for polyphonic music. For purposes of deriving transforms, note sequences were extracted by removing all aspects of the repre-

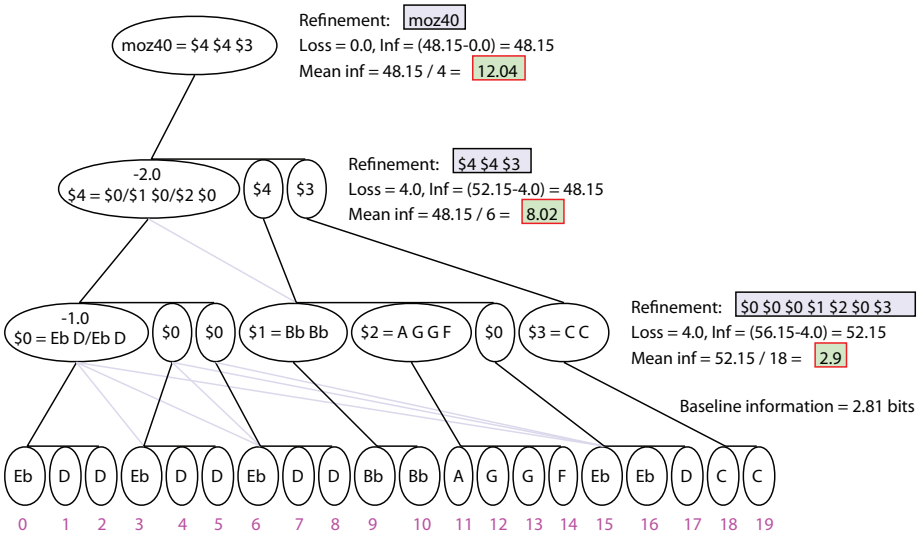

Figure 4: Bayes transform of pitches from the theme of K.500.

sentation except for the note-codes themselves, and by collapsing tied notes to single values, as in (Pearce *et al.* 2008).

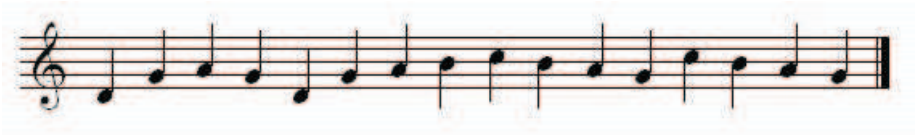

Figure 5: Melody 'czech30' from the 'czech' subset of the Essen corpus.

Note codes in **kern format commence with an integer indicating note duration. The value 1 represents a whole note, 2 represents a half-note, 4 represents a quarter note, and so on. Dots may be added to increase duration by 1/2. Appended immediately after the duration comes the pitch. This is represented as an alphabetical character, with an optional '#' to represent a semitone increase, or '-' to represent a semitone decrease. Lower-case values represent pitches in the octave immediately above middle C, while a pitch in an octave above is indicated by repeating the character the relevant number of times. Lower-case characters are then used correspondingly to represent octaves below. (Documentation on the **kern representation can be found at the website kern.ccarh.org).

Consider the melody 'czech30' from the 'czech' subset (Figure 5). Represented using **kern note codes, this becomes the sequence '4d 4g 4a 4g 4d 4g 4a 4b 4cc 4b 4a 4g 4cc 4b 4a 4g', from which we obtain the transform of Figure 6. The suggested phrasing for 'czech30' comprises one (internal) phrase bound-

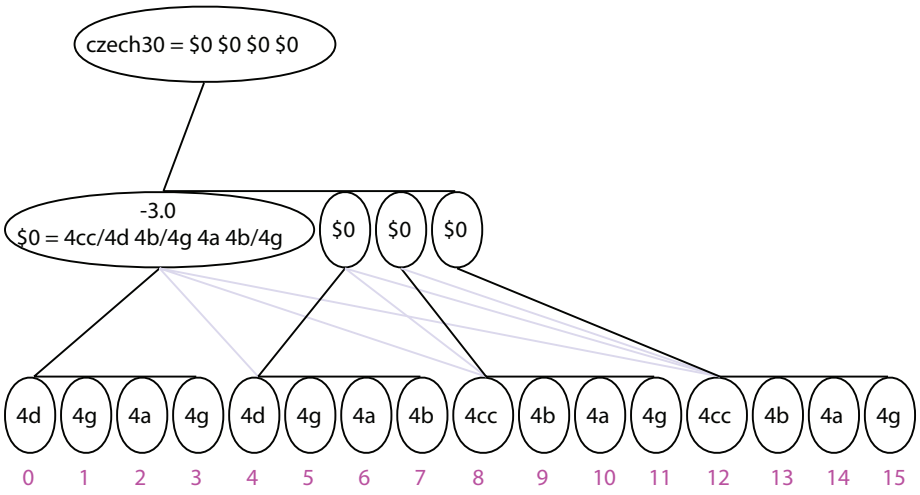

Figure 6: Bayes transform of 'czech30'.

ary, occurring after the 7th event in the sequence (the quarter note '4b'). The phrasing of the piece thus consists of one phrase comprising events 0-7, and another comprising events 8-15. Looking at the transform of Figure 6, we see that the first refinement breaks the data up in the same way, although dividing the two specified phrases into two sub-groups. The correspondence between the transform and the (corpus-specified) phrasing is accounted for by the degree to which phrasing in this melody is mediated by repetition.

To evaluate the degree to which phrasing structures correspond to transform structures in general, Bayes transforms for all songs in the corpus were derived, with measurements being made of the degree of correspondence in each case. For each refinement level of each transform, a calculation was made of the number of cases in which a specified phrase boundary was seen to be identical to a construct boundary. The highest value obtained was then used to determine a *recall* value (Manning and Schütze, 1999), calculated as the number of aligned boundaries expressed as a proportion of the total number of specified boundaries. Calculating recall values in this way, we then see how different subsets of the corpus show different degrees of transform/phrasing alignment. Recall values were found to be in the range 0.55-0.87 over the wholse corpus, with a mean of The mean recall over the whole corpus was 69.7%. This suggests Essen melodies typically exhibit a fair degree of correspondence between phrasing and repetition hierarchy, as might be expected with this musical style.

Analysis of phrasing in the Essen corpus has generally aimed to produce models that can predict the placement of boundaries. Transforms are not straightforwardly applied to this task, however: they typically represent hierarchical structure at several levels, whereas the specification in the corpus represents structure at a single level. Predicting phrasing on the basis of a

transform then begs the question of which refinement to treat as a reference.

We can obtain basic prediction protocols simply by fixing the decision in advance, however. Testing data that exhibit transform/phrasing correspondence at a specific level, we then obtain the prediction accuracies of Table 1. Each row, here, represents accuracies obtained by basing predictions on a particular refinement level. The scores for 'predict-1' were obtained by predicting phrase boundaries to be aligned with first-refinement constructs, 'predict-2' by predicting phrases to align with second-refinement constructs, and so on. The predict-1 strategy produces a recall of nearly 100%. Unfortunately, this is associated with a precision of only 14%. Predict-1 produces few false negatives but an extremely high proportion of false positives.

Strategy	Precision	Recall	F1 score
predict-1	0.14	0.99	0.24
predict-2	0.16	0.47	0.24
predict-3	0.12	0.14	0.12

Table 1: Phrase-prediction scores using different refinement levels

Various results have been reported for the task of predicting phrasing in the Essen corpus. A study by Temperley (2001) reported an accuracy of 75.5%, using the preference-rule method *Grouper*. This was a measure of recall derived using a subset of 65 cases — just over 1% of the corpus. This subset was selected by choosing four cases from each region and then deleting those (15) cases which exhibited metrical irregularities. *Grouper*'s 75.5% recall accuracy was compared by Bod (2002b) to the 87.3% figure he obtained using a parse-based method (discussed at more length below). However, Bod's figure was derived by predicting phrase groups rather than internal boundaries. It was also an $F1$ score rather than a percentage, derived by combining precision and recall values in the formula

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Other results for this task have been reported by (Pearce *et al.* 2008; Pearce *et al.* 2010). This study examined the 1700 songs from the 'deutschl/erk' subset, which forms nearly a quarter of the entire corpus. This showed *Grouper* to produce a recall of 62%, somewhat lower than the 75.5% Temperley achieved using the 65-song subset. This study also tested the IDyOM method, discussed below; this was shown to produce a recall of 50%.

Comparing these results with the scores of Table 1 is complicated, since all figures were derived using different subsets of the corpus. Furthermore, some values (such as the recall figures for IDyOM and Grouper) were derived in a fully unsupervised way, while others (e.g., the $F1$ figure for Bod's method) were derived using a supervised learning regime that involved sampling all the available (non-test) data in the corpus.

Arguably, the performance of the best performing transform-based method 'predict-1' — which predicts phrases to align with constructs of the 1st refinement — is most directly comparable with that of IDyOM. Both methods are fully unsupervised and both make no use of domain knowledge. IDyOM's recall performance on the 'erk' songs was reported to be 60%, while predict-1 produced a recall of 99% on a different but overlapping selection of cases. However, the precision performance of 0.14 obtained for predict-1 is significantly worse than the 0.76 reported for IDyOM on the 'erk' songs. The situation with the $F1$ scores is similar. IDyOM achieves an $F1$ score of 0.58 on the 'erk' songs, whereas predict-1 produces 0.24 on the corpus as a whole.

3.2 Generative use of transforms

Where Bayes transforms are taken from sequential data (and constructs are subsequences), refinements necessarily maintain the ordering of entities seen in the original data. The constituents of constructs at any level of refinement then reflect the left-to-right ordering of symbols in the sub-dataset. A complete refinement hierarchy has the 'grammatical' structure of an AND/OR tree. In this respect, transforms are akin to stochastic phrase-structure grammars, which augment an ordinary PSG with rule-related probabilities. Being implicit phrase-structure grammars, Bayes transforms can be represented in the form of trees, as below and elsewhere. They are also straightforwardly deployed for generative purposes, via the procedure of top-down, symbol expansion.

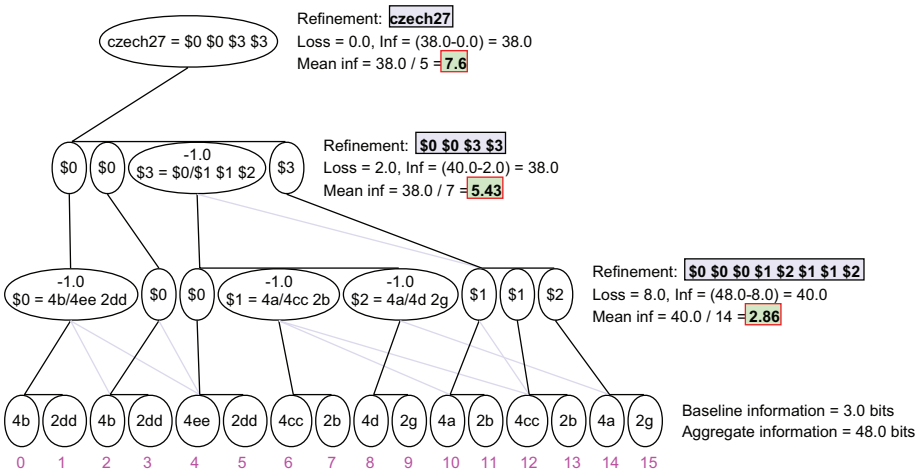

Figure 7: Transform of 'czech27'.

Applied to a phrase-structure grammar, this procedure has a simple, recursive form (Lerdahl and Jackendoff, 1983). Beginning with the start symbol of the grammar, we repeatedly replace symbols with their constituents, making

random choices where alternatives are defined. Continuing on until no unexpanded symbols remain, we eventually obtain a complete instance derived from the grammar. Applied to a transform, the procedure works in much the same way, but commencing with expansion of the root. The process can be illustrated using the example of the 'czech27' melody from the Essen corpus. Comprising just 15 notes, this is one of the shortest melodies in the 'czech' subset. Its transform appears in Figure 7.

Consider application of symbol-expansion to this transform. First we expand the root symbol 'czech27' by replacing it with its four constituents '$0 $0 $3 $3'. We then seek to expand each of these symbols in the same way. $3 is a generalizing construct, however, allowing either 4a or 4ee as its initial constituent. It may be expanded to form either '4a $1 $2' or '4ee $1 $2': we randomly select one of those forms. Continuing on, we eventually arrive at a sequence of primitive note values. Treating the transform as a structural model of the original dataset, this derived sequence can be viewed as a structural/statistical variant of the original melody.

Depending on how choices are resolved, different instances may be derived. To obtain the entire set of possible derivations, we must pursue all possibilities to a final conclusion in every case. Applying this process expansion to the case of 'czech27', we obtain a set of 64 derivations, three of which are shown in Figure 8.

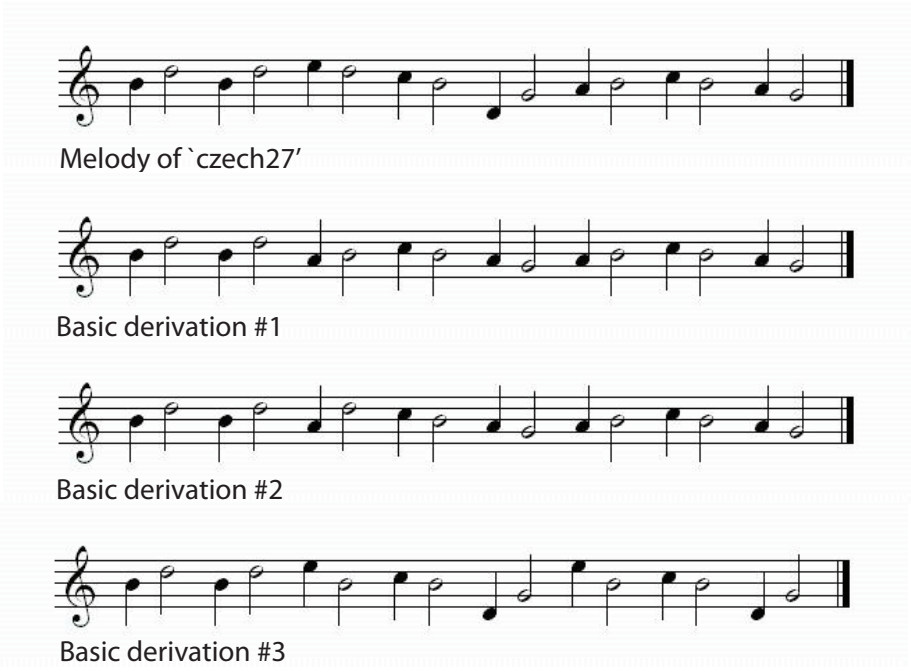

Figure 8: Three of the 64 'basic derivations' from czech27's transform.

In this generative regime, instances are derived by applying symbol elaboration to the root of a single melody's full transform. Termed 'basic derivation' below, this is the simplest way of using transforms generatively. Its capacity to produce musically interpretable results depends on the original material. Where musical properties are reflected in repetition structures, there is the potential for these to be implicitly modeled in the transform. Derivation of instances may then yield results that re-express those properties in a new way. The melody of 'czech27' is an advantageous case in this respect. The derived transform implicitly models certain properties of phrasing; these are then re-expressed in derived instances, albeit with results that do not differ significantly from the original melody.

A practical difficulty with basic derivation is the number of instances that can typically be obtained. Illustrating the effect is the case of 'romani18'. While not a particularly complex melody, this yields a transform of four refinement levels, with generalization at several levels of organization, as shown in Figure 9. The greater degree to which the transform models repetition-hierarchy promises more robust inheritance of properties. But we now face the difficulty of vast generative capacity. Due to the greater use of generalization, there are in fact more than 10^{16} instances that can be derived from this transform.

3.3 Controlled derivation

A less profligate generative regime is that of *controlled* derivation. This focuses on the states of a particular construct in the transform. The essence of the procedure is to take each state (i.e., each choice combination) of the construct in turn, and derive an instance in the usual way, letting all other constructs take whatever state serves to minimize similarity with the original sequence. (For this purpose, similarity is defined to be the proportion of cases where the same element appears in the same position in the sequence.) This typically produces n derivations for a construct with n states, although there may be more if there are multiple derivations which are all maximally similar to the original sequence. As a simple illustration, Figure 10 shows the transform of 'romani10', with Figure 11 contrasting the original melody against two of the eight instances derivable from the second \$2 construct in the second refinement of the transform.

3.4 Composed derivations

When instances are derived from transforms using basic or controlled derivation, symbol-elaboration commences with the root of a single transform. The resulting instance then forms a statistical variant of the data from which the transform was derived. Any construct can be deployed as the starting point for symbol-elaboration, however: it need not be a root. Indeed, the starting point can be any *sequence* we like. In this way, we can obtain transforms using seeds taken from from different levels of the same transform, or from different transforms altogether.

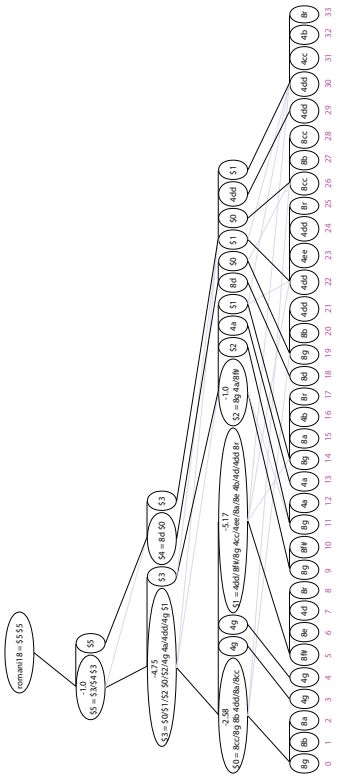

Figure 9: Transform of ‘romani18’.

Exploiting these possibilities produces the protocol of *composed* derivation. Symbols drawn from one or more refinement levels (in one or more transforms) are combined to form an initial sequence. This then becomes the generative schema seeding the symbol-expansion process. Instances so derived have the potential to reflect aspects of repetition-hierarchy originating in different melodies. Most simply, this approach provides the means of combining passages from different melodies in a ‘cut-and-paste’ style of composition. But where symbols at higher levels of refinement are deployed, each one defines a set of possible outcomes. A derived instance then consolidates a combination of constituents drawn from sets of possibilities.

This protocol allows a much greater degree of user choice than either basic or controlled derivation. The results obtained may depend rather significantly on how effectively this choice is exercised. As an illustration of what is generally achieved, consider Figure 13. This shows four instances derived from the schema ‘$17, $3, $17, $10’. This is a sequence of symbols drawn from three different transforms. Symbol $17 represents a construct from the transform of ‘jugos004’. Symbol $3 represents a construct from the transform

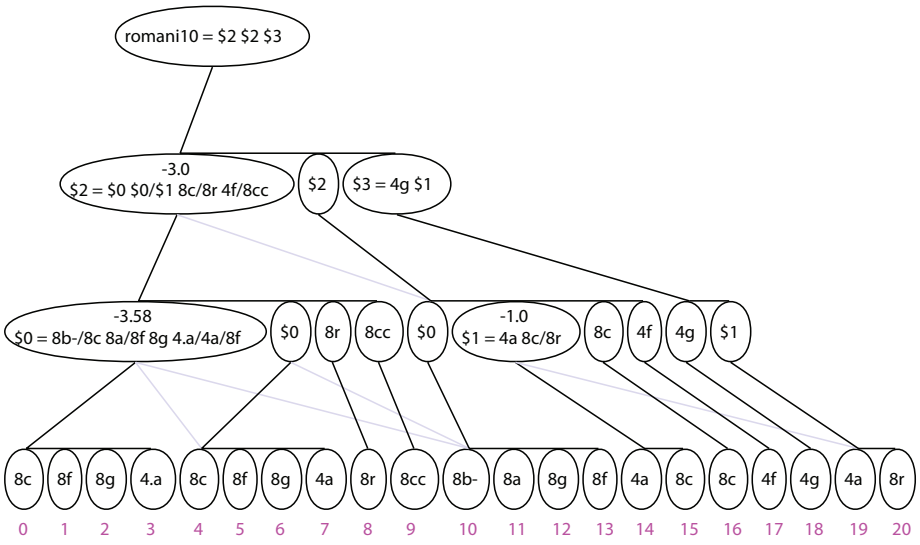

Figure 10: Transform of 'romani10'.

of 'czech08', while symbol '$10' represents a construct from the transform of 'ukrain12'. (Scores for these melodies can be viewed at the accompanying webpage www.ChrisThornton.eu/folksong-transforms.html). Combined in a single schema, these then serve as seeds for different avenues of generative symbol-expansion. Derived instances then embody statistical features of data based on four, original melodies.

3.5 The problem of *inter opus* patterns

Examples presented so far all use transforms extracted from single melodies. When this approach is taken, only those patterns of repetition existing within the relevant melody have any impact. Factors relating to musical style may involve commonalities between relevant pieces, as much as within them, however. Generative methods are ideally applicable to sets of melodies, enabling exploitation of *inter opus* patterns.

Since the Bayes transform has the potential to be applied to datasets of any structure, a possible approach to this issue would be to devise a specialized data representation for musical corpora (cf. Cope, 2001: Cope, 2005). Generative protocols appropriate to that representation could then be devised. More straightforwardly, there is the possibility of forming aggregated datasets of the sequential type, simply by chaining melodies together. We might form a chain of all 28 melodies in the 'romania' subset, for example. The transform then derived will model the repetition hierarchy of the chain. Where commonalities exist between individual melodies, these will promote constructs that generalize

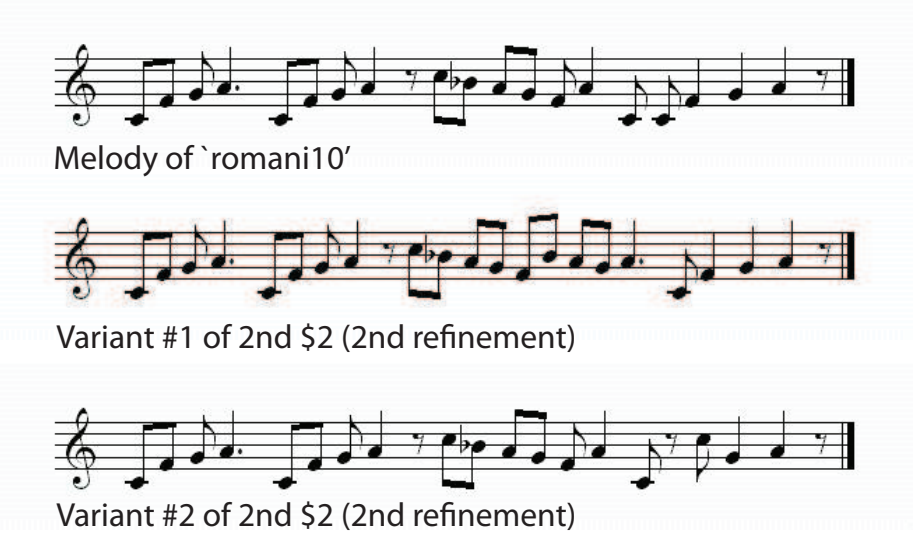

Figure 11: Melody 'romani10' with two of eight derivations based on construct $2. Notice the first 8 notes of each melody are the same.

material found in multiple melodies. Extracting that sub-tree of the complete transform that originates in a specific melody, we then obtain a structure that may reflect *inter opus* patterns.

An obvious problem with this approach is its computational cost. With a greater number of melodies in the chain, the computational effort required to obtain the transform increases. In practice, this does not seem to be a problem for modest numbers of melodies, however. It has been possible to process all 28 melodies from the 'romania' subset, for example, in less than three minutes on a normal PC.

As an illustration of this approach, consider Figure 13. This shows the sub-transform for melody 'sverig05', extracted from the transform of the chain comprising all melodies from the 'sverige' subset. Construct specifications show only those constituents within the 'sverig05' section of the chain, and thus offer no overt evidence of *inter opus* patterns. However, the way in which the constructs divide up the data is mediated by the informational properties of the chain. To that extent, refinements take account of repetition bridging multiple melodies.

Illustrating the generative possibilities of the approach, the score of Figure 14 is derived from the single variant of construct $65, taken from the 'sverig05' sub-transform.

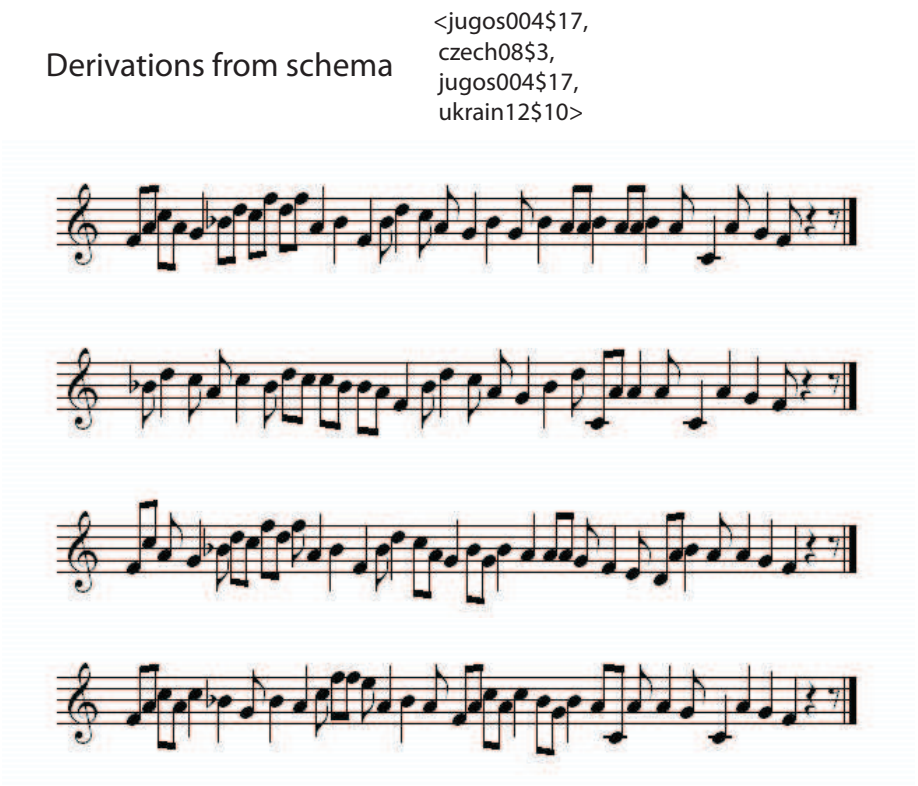

Figure 12: Composed derivation using constructs from three transforms.

4 Discussion

The aim of the paper has been to see what can be achieved by treating the Bayes transform as a way of modeling hierarchical structure in melodic sequences. It bears repeating that the method is not proposed to be a new form of music analysis. Motivated on purely informational grounds, the procedure can be used to place sequential data of *any type* into an efficient, hierarchical representation. It is as easily applied to sequences of words (or numbers) as it is to sequences of notes.

Any virtue the procedure may have for modeling music arises from its ability to capture patterns of repetition (and approximate repetition) at different levels of organization. Where hierarchical patterns of repetition are musically meaningful, the Bayes transform may then have some musical salience, and generative results some musical properties. Whether they do so in practice seems largely a question of subjective evaluation. To aid in the formation of evaluations, a public web-page has been provided (at www.ChrisThornton.eu/folksong-transforms.html), from which a large number of generative examples can be assessed (including

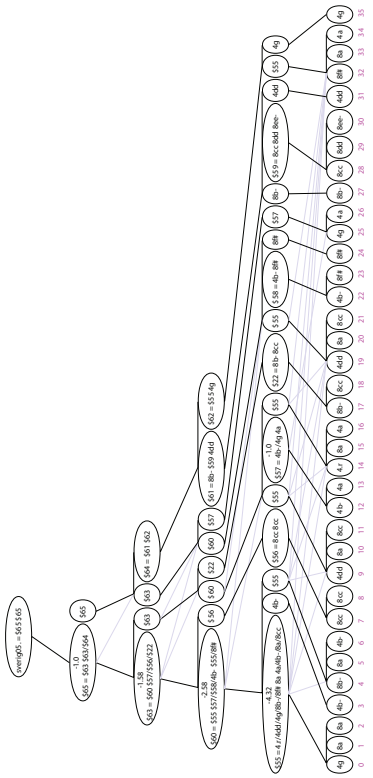

Figure 13: Sub-transform extracted from transform for 'sverige' chain.

all examples above), both in score form, and through audio playback.

To the extent that these generated note sequences are judged to have musical credibility, the question then arises as to *why* this should be the case. Intuitions about how Bayes transforms capture patterns of repetition may form part of the answer. But, is there anything more substantive to be said? Specifically, is there any way in which Bayes transforms reflect, or express theoretically informed lines of analysis?

It is certainly the case that repetition is widely recognized to be salient in musical meaning (Brown and Dempster, 1989). Methods for *detecting* repetition often have features in common with the Bayes transform. Consider research reported by (Meredith *et al.* 2002). Proceeding from the observation that 'there are certain types of interesting musical repetitions that cannot be discovered using string algorithms' (ibid., p. 321), Meredith *et al.* propose the algorithms SIA and SIATEC for discovery and representation of 'patterns with gaps'. As in the present proposal, the approach involves deployment of higher-level representations. While these are specifically geometric in character, their role is to accommodate the way 'patterns involved in ... repetitions [may] vary widely in

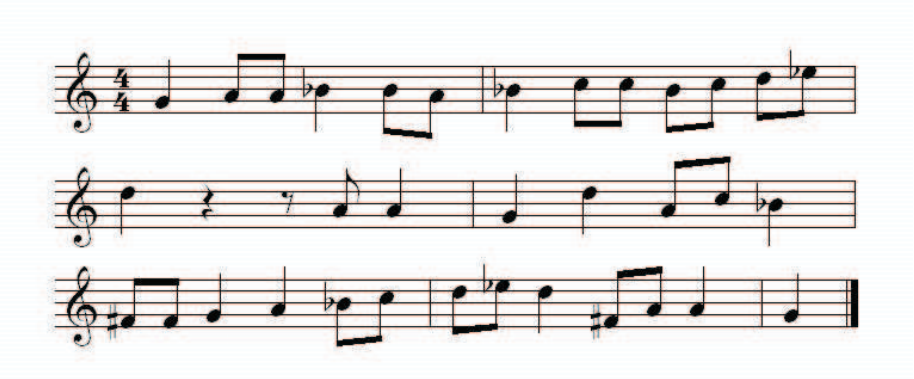

Figure 14: Variant of $65 (sverige chain-transform extract for sverig05).

their structural characteristics' (ibid., p. 322). Emphasising structural aspects of repetition, Meredith et al. also stress the significance of hierarchy. As they note, a 'perceptually significant repeated pattern may be a small motive consisting of just a few notes... or it might be a whole section of the work' (ibid., p. 322).

The present proposal shares Meredith et al.'s general strategy of deploying higher-level representation for capture of repetition structure. But while Meredith et al. observe that 'compactness is one important feature that is common to most theme-like patterns' (ibid., p. 341), their methods do not directly exploit representational simplicity. This may be part of the reason SIA/SIATEC cannot exploit patterns of approximate repetition. Methods for that task (e.g., Cambouropoulos *et al.* 1999; Rolland, 1999) are noted to typically involve measurement of *edit distance* — roughly, the number of changes that have to be made to transform one sequence into another. Given constant-length comparisons, this is also an implicit feature of the Bayes transform. By minimizing aggregate information loss, generalization automatically favours constructs with maximum material in common.

In other respects, the Bayes transform departs more significantly from the repetition-discovery paradigm. The goal of the procedure is not discovery of patterns. Rather, it is production of a representation which efficiently encodes the original sequence in terms of repeating patterns. This links the approach more strongly with work on production of summary representations.

Relevant, here, is work using Self-Organizing Maps (Kohonen, 1984) as a representational medium. Kohonen himself has developed a data-driven approach for generative music, based on two phases of operation (Kohonen, 1989). In the first phase, musical sources are scanned to determine the set of context rules that will unambiguously determine the continuation of notes from each point. (Kohonen terms this the *Dynamically Expanding Context* or DEC method.) In the second phase, the SOM learning procedure (Kohonen, 1988b) is used to

obtain a low-dimensional representation (i.a., a map) for the context rules derived. Much in the manner of the 'phonetic typewriter' (Kohonen, 1988a), it is then possible to classify existing forms in terms of points in the map. There is also the possibility of generating new musical forms by inscribing points or trajectories into the map (Kohonen *et al.* 1991).

Information refinement proceeds along similar lines. It also yields a summary representation based on patterns found in relevant sources. But there is no commitment to a topographical form of encoding. And the utilization of relatively short patterns emerges from optimization of information rather than from explicit rule-derivation. Kohonen's method may not directly apply information-theoretic principles. But it can be seen to do so implicitly. The attempt to determine rules which unambiguously predict continuations on the basis of minimum preceding context can be viewed as the attempt to minimize uncertainty about how the music extends from any given point.

Interpreting Kohonen's rule-derivation in this way helps link it to work that makes more explicit use of information theory for prediction of phrasing. Approaches such as (Juhász, 2004; Pearce *et al.* 2008, Juhász, 2009) derive information about boundaries using explicit measurements of uncertainty (i.e., entropy). These methods treat musical events (e.g., notes) as a stream of signals. Depending how the music arrives at a particular point, the way in which it continues then has a certain level of predictability. Less predictability implies greater uncertainty. By calculating the predictability of continuations, it is possible to establish an uncertainty contour for the piece. Peaks and troughs in this contour can then provide cues about boundaries in the piece.

This approach has been used with considerable success for prediction of phrase boundaries in the Essen corpus (e.g., Pearce and Müllensiefen, 2008; Juhász, 2009). But the way the approach is applied can make quite a difference. There is the difficult question of how much preceding context to take into account when calculating predictions. There is also the question of how the preceding context should be represented to bring out salient musical properties.

The question of how much context should be taken into account is challenging in itself. As Conklin and Witten note, the difficulty is that 'very low order models are too general, and do not capture enough structure of the concept; very high order models are too specialized to the examples from which they were constructed, and do not capture enough statistics of the concept' (Conklin and Witten, 1995, p. 56-57). (Here, the term 'order' identifies the number of preceding entities assumed to comprise the context.) Conklin and Witten's approach overcomes the problem by deploying multiple representations simultaneously. Predictions of future note events are generated by weighting distributions from separate representations, such that 'viewpoints that are very uncertain about the outcome are given lower weight' (Conklin and Witten, 1995, p. 61).

The question of how context should be represented is equally problematic. Representing context in terms of explicit note values is generally deemed overspecific, since the same motive may arrange notes in slightly different ways. The solution seems to be to deploy more abstract representation. But there are a range of ways in which this can be done. In Conklin and Witten's (1995)

framework, predictions about future note events are based on structures which represent the preceding context in terms of musical 'viewpoints'. This level of representation abstracts away small-scale differences in motives, enabling more effective deployment of uncertainty measurement.

In the approach of Juhász (Juhász, 2002; Juhász, 2004; Juhász, 2009), map-like representations are derived for 'motive contours'. Uncertainties calculated in terms of locations in the map then become the means of predicting boundaries. Juhász has examined a range of ways in which topographical representations can be obtained, including principle components analysis (Juhász, 2002) and Kohonen's SOM learning method (Juhász, 2009). Others follow similarly motivated approaches (e.g., Toiviainen and Eerola, 2002).

The Bayes transform has elements in common with these methods. It shares with them the assumption that any musical piece exhibits a contour of informational uncertainty. It seeks to use this information in much the same way, i.e., for identifying internal boundaries. But it differs in its procedure. In a method such as (Juhász, 2009), derivation of the summary representation, and derivation of the segmentation are separate steps. Moreover, informational considerations deployed in the second step do not affect the outcome of the first. In the Bayes transform, the two steps are unified. Segmentation emerges implicitly through summarization. Or we can see it *vice versa*. The process of deriving a summary representation *is* the process of predicting segmentation boundaries.

Other differences are also worth highlighting. In the Bayes transform, the summary representation obtained is intrinsically hierarchical in nature. It is the *nesting* of this structure that reflects the properties of relevant information contours. A slightly more general assumption is then in force. On the basis that a piece will tend to exhibit informational contours at different levels of organization, hierarchical descriptions obtained (from the Bayes transform) reflect the way in which an informational contour at one level is nested within in an informational contour at another.

The Bayes transform is thus sensitive not just to information at one level or organization. It is sensitive to information at multiple levels of organization. The hierarchical representation obtained reflects the way uncertainty contours are nested one within another. This structure-orientation more strongly reflects the parse-tree approach of Bod (2002b) to the problem of predicting phrase boundaries. In Bod's approach, there is no direct calculation of continuation uncertainties, as in methods such as (Juhász, 2004). Rather, the goal is to construct a model in terms of parse trees that best describe the source material. Candidate trees are weighted according to the frequency with which their instances occur in the source material. Bod found that the best model overall is obtained by combining a likelihood preference (favouring models of higher probability) with a simplicity preference (favouring simpler models).

While this process may not apply information-theoretic measurement explicitly, it can be viewed as doing so implicitly. As Temperley (2004) observes, production of a representation which makes observed data a relatively more probable prediction from a relatively smaller number of factors can be formalized in Bayesian terms as minimization of cross-entropy between model and data

(Temperley, 2007; Temperley, 2008). Bod's simplicity/likelihood heuristics can then be seen as applying an information-theoretic preference for uncertainty-minimizing representations of a hierarchical nature. This further emphasizes the link between Bod's approach and the Bayes transform.

Indeed, the Bayes transform can be seen as addressing a problem noted with Bod's approach, relating to structure of rules. As originally formulated, Bod's (2002a) method relies exclusively on non-disjunctive parse trees. This is seen as a problem by Temperley, who notes 'the model "learns" about phrase structure from examples, but does not appear to generalize.' (Temperley, 2007, p. 147). The concern here is that generalization in structural representation of music is generally taken to be a necessity (Temperley, 2001; Longuet-Higgins, 1976; Tenney and Polansky, 1980).

In the structural representations that emerge under the Bayes transform, uncertainty is traded for compactness in a way that optimizes mean information at each level. Introduction of uncertainty implies introduction of choice, i.e., creation of disjunction. The structural rules expressed by a Bayes transform then accommodate the properties of generalization we associate with grammatical representation. While having some similarity with Bod's parse-tree approach, the Bayes transform thus also links with more specifically *grammatical* approaches to representation.

The point of contact then becomes Lerdahl and Jackendoff's 'Generative Theory of Tonal Music' (1983) (henceforth the 'GTTM'). This theory also makes no direct use of information-theoretic principles. It is oriented towards provision of well-formedness and preference rules for describing musical constructs. The grouping well-formedness rule 'GWFR 1', for example, states that 'Any contiguous sequence of pitch-events, drum beats, or the like can constitute a group, and only contiguous sequences can constitute a group.' (ibid., p. 37). The generative power of the framework derives from deployment of these representations in grammar-like, hierarchical structures.

Rather than committing to an approach which focuses on motives at one level of description, Lerdahl and Jackendoff emphasize the way patterns exist at multiple levels. As they comment, 'An obvious observation about music is that some musical passages are heard as ornamented versions, or *elaborations* of others' (ibid., p. 105). Certain passages are 'heard as elaborations of an abstract structure that is never overtly stated' (ibid., p. 105), but which is a 'simplified' form of the surface manifestation. These observations then lead to the Reduction Hypothesis, under which the listener is proposed to 'organize all the pitch events of a piece into a single coherent structure, such that they are heard in a hierarchy of relative importance' (ibid., p. 105).

In Lerdahl and Jackendoff's view, this concept of a multi-level, grammar-like structure most naturally expresses a Schenkerian theory of music (Schenker, 1935/1979). But for present purposes it can also be seen as reworking Bod's notions of information-rich hierarchical description, in the general context of information refinement. The ability of Bayes transforms to capture generalizations at multiple levels of representation means they are fully able to express the grammatical structures of the GTTM. But the informational basis of the derivation

means we also have the possibility of treating the emergence of a GTTM-style representation as directly implementing minimization of cross-entropy between model and data, in the manner emphasized by Temperley (2007).

Linking the GTTM with information theory in this way may seem contrary to the author's intentions. Lerdahl and Jackendoff explicitly argued that mathematically-based analysis is inappropriate for representation of music. The fact that mathematics is 'capable of describing any conceivable type of organization' (ibid., p. 2), they argue, indicates that it lacks the requisite element of selectivity. 'To establish the basis for a theory of music,' they assert, 'one would want to explain why certain conceivable constructs are utilized and others not' (ibid., p. 2).

But Lerdahl and Jackendoff's position on this point may reflect a Chomskyan assumption (Chomsky, 1957), in which mathematical optimization is held to be fundamentally discontinuous with emergence of structural representation. Under the present proposal, this discontinuity is eliminated. Processes of mathematical optimization become the means of producing structural forms that have the generative power of Chomskyan grammars. The approach then provides an explanation why 'certain conceivable constructs' are preferred in modeling certain musical forms. More significantly, the explanation is expressed in terms of the general principles of information theory.

Summing up, the Bayes transform *can* be shown to reflect theoretically-informed music analysis in a number of ways. Broadly Bayesian connections between seemingly divergent approaches begin to become more apparent, while the recursive refinement model of Section 3 allows us to see where common assumptions about information usage are being made. There then seems to be less distance between structurally oriented approaches (such as the GTTM) and statistically oriented approaches, such as (Pearce and Müllensiefen, 2008) and (Juhász, 2009). Treating the Bayes transform as a structural mediation for Temperley's (2007) entropy-minimization criterion, it is reasonable to infer a certain underlying continuity between them.

Appendix 1: Repetition-phrased songs

The following listing presents the names of all 183 songs from the Essen corpus that exhibit perfect phrasing/transform correspondence, i.e., have at least one refinement level producing a recall performance of 100% with a precision of at least 20%.

czech06 deut3711 deut3777 deut3840 deut3865 deut3883 deut3937 deut4014 deut4101 deut4117 deut4154 deut4224 deut4289 deut4313 deut4337 deut2999 deut3258 deut3455 deut3456 deut3554 deut3573 deut3616 deut2333 deut2598 deut2616 deut2637 deut2655 deut2742 deut2796 deut2819 deut2827 deut2893 deut2922 deut2959 deut4406 deut4412 deut4451 deut4483 deut0627 deut0830 deut0834 deut0839 deut0848 deut0916 deut0954 deut0993 deut1066 deut1071 deut1080 deut1166 deut1191 deut1230 deut1278 deut1282 deut1318 deut1378 deut1385 deut1416 deut1490 deut1505 deut1539 deut1545 deut1547 deut1575

deut1577 deut1593 deut1610 deut1638 deut1641 deut1650 deut1655 deut1685 deut1693 deut1700 deut1716 deut1729 deut1762 deut1796 deut1831 deut1858 deut1874 deut1914 deut1919 deut1946 deut1975 deut1982 deut1991 deut1995 deut2023 deut2061 deut2070 deut2089 deut2090 deut2098 deut2100 deut2103 deut2104 deut2113 deut2124 deut2160 deut2165 deut2265 deut025 deut029 deut035 deut074 deut079 deut082 deut122 deut207 deut381 kindr001 kindr005 kindr009 kindr032 kindr033 kindr039 kindr046 kindr051 kindr058 kindr069 kindr076 kindr079 kindr094 kindr095 kindr100 kindr111 kindr113 kindr116 kindr126 kindr131 kindr143 kindr148 kindr151 kindr163 kindr167 kindr170 kindr171 kindr181 kindr189 kindr196 kindr201 kindr209 kindr210 deut4603 deut4617 deut4627 deut4639 deut4658 deut4676 deut4684 deut4723 deut4761 deut4774 deut4807 deut4876 deut4905 deut4945 deut5004 deut5036 deut5040 deut5046 deut5057 deut5087 elsass59 elsass72 elsass76 elsass85 england3 jugos018 jugos035 lothr025 magyar31 neder026 neder053 oestr030 oestr080 oestr096 polska15 polska20 romani02 suisse21 ukrain05

References

Bent, I. and Drabkin, W. (1987). *Analysis: New Grove Handbooks in Music*, Macmillan.

Bergeron, M. and Conklin, D. (2007). Representation and discovery of feature set patterns in music. *Proceedings of the International Workshop on Artificial Intelligence and Music: 20th International Joint Conference on Artificial Intelligence* (IJCAI) (pp. 1-12). Hyderabad, India.

Blumer, A., Ehrenfeucht, A., Haussler, D. and Warmuth, M. (1987). Occam's razor. *Information Processing Letters*, *24* (pp. 377-380).

Bod, R. (2001). Probabilistic grammars for music. *Belgian-Dutch Conference on Artificial Intelligence*.

Bod, R. (2002a). Memory-based models of melodic analysis: challenging the gestalt principles. *Journal of New Music Research*, *31*, No. 1 (pp. 27-36).

Bod, R. (2002b). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, *17* (pp. 289-308).

Brown, M. and Dempster, D. (1989). The scientific image of music theory. *Journal of Music Theory*, *33*, No. 1.

Cambouropoulos, E., Crochemore, M., Iliopoulos, C., Mouchard, L. and Pinzon, Y. (1999). Algorithms for computing approximate repetitions in musical sequences. In R. Raman and J. Simpson (Eds.), *Proceedings of the 10th Australasian Workshop on Combinatorial Algorithms* (AWOCA-99) (pp. 129-144). Perth, WA, Australia.

Chaitin, G. (1979). Toward a mathematical definition of life. In I.R. Levine and M. Tribus (Eds.), *The Maximum Entropy Formalism* (pp. 477-498). MIT Press,.

Chater, N. (1999). The search for simplicity: a fundamental cognitive principle?. *The Quarterly Journal of Experimental Psychology,*, *52A*, No. 2 (pp. 273-302).

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Cohen, J. (1962). Information theory and music. *Behavioral Science*, *7* (pp. 137-163).

Conklin, D. (2002). Representation and discovery of vertical patterns in music. In C. Anagnostopoulou, M. Ferrand and A. Smaill (Eds.), *Music and Artificial Intelligence: Proceedings ICMAI 2002* (pp. 32-42). Springer-Verlag,.

Conklin, D. (2006). Melodic analysis with segment classes. *Machine Learning*, *65* (pp. 349-360).

Conklin, D. and Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *INFORMS Journal on Computing*, *18*, No. 3.

Conklin, D. and Witten, I. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, No. Issue 1 (pp. 51-73).

Cope, D. (2001). *Virtual Music: Computer Synthesis of Musical Style*. London: The MIT Press.

Cope, D. (2005). *Computer Models of Musical Creativity*. The MIT Press.

Dannenberg, R. and Hu, N. (2003). Pattern discovery techniques for music audio. *Journal of New Music Research*, *32*, No. 2 (pp. 153-163).

Held, G. (1987). *Data Compression: Techniques and Applications, Hardware and Software Considerations* (2nd edition). Chichester: Wiley.

Hiller, L. and Fuller, R. (1967). Structure and information in webern's symphonie, opus 21. *Journal of Music Theory*, *11* (pp. 60-115).

Hillewaere, R., Manderick, B. and Conklin, D. (2009). Global feature versus event models for folk song classification. *10th International Society for Music Information Retrieval Conference* (ISMIR 2009) (pp. 729-733).

Huron, D. (1999). *Music Research Using Humdrum: A User's Guide*. Stanford, Calif.: Center for Computer-Assisted Research in the Humanities: http://dactyl.som.ohio-state.edu/Humdrum/guide.toc.ht ml.

Juhász, Z. (2002). The structure of an oral tradition - mapping of hungarian folk music to a metric space. *Journal of New Music Research*, *31*, No. 4 (pp. 295-310).

Juhász, Z. (2004). Segmentation of hungarian folk songs using an entropy-based learning system. *Journal of New Music Research*, *33*, No. 1 (pp. 5-15).

Juhász, Z. (2009). Automatic segmentation and comparative study of motives in eleven folk song collections using self-organizing maps and multidimensional mapping. *Journal of New Music Research*, *38*, No. 1 (pp. 71-85).

Knopoff, L. and Hutchinson, W. (1981). Information theory for musical continua. *Journal of Music Theory*, *25*, No. 1 (pp. 17-44).

Knopoff, L. and Hutchinson, W. (1983). Entropy as a measure of style: the influence of sample length. *Journal of Music Theory*, *27* (pp. 75-97).

Kohonen, T. (1984). *Self-organization and Associative Memory*. Berlin: Springer-Verlag.

Kohonen, T. (1988a). The 'neural' phonetic typewriter. *Computer*, *21*, No. 3 (pp. 11-22).

Kohonen, T. (1988b). *Self-organization and Associative Memory* (second edition). New York: Springer-Verlag.

Kohonen, T. (1989). A self-learning musical grammar, or "associative memory of the second kind". *Proceedings of the 1989 International Joint Conference on Neural Networks*. San Diego.

Kohonen, T., Laine, P., Tiits, K. and Torkkola, K. (1991). A nonheuristic automatic composing method. In P.M. Todd and G. Loy (Eds.), *Music and Connectionism* (pp. 229-240). Cambridge, MA: The MIT Press.

Lerdahl, A. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT Press.

Li, M. and Vitányi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications: Second Edition*. New York: Springer-Verlag.

Longuet-Higgins, H. (1976). Perception of melodies. *Nature*, *263* (p. 646.653).

Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Margulis, E. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal*, *32*, No. 4.

Meredith, D., Lemstrvm, K. and Wiggins, G. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, *31*, No. 4 (pp. 321-345).

Meyer, L. (1957/1967). Meaning in music and information theory. *Music, the Arts, and Ideas* (pp. 5-21). Chicago: University of Chicago Press.

Pearce, M., Conklin, D. and Wiggins, G. (2005). In U. Wiil (Ed.), *Methods for Combining Statistical Models of Music*. Computer Music Modelling and Retrieval (pp. 295-312). Heidelberg, Germany: Springer Verlag.

Pearce, M., Müllensiefen, D. and Wiggins, G. (2008). *An Information-dynamic Model of Melodic Segmentation*. Helsinki, Finland: International Workshop on Music and Machine Learning.

Pearce, M., Müllensiefen, D. and Wiggins, G. (2008). A comparison of statistical and rule-based models of melodic segmentation. In J.P. Bello, E. Chew and D. Turnbull (Eds.), *Proceedings of the 9th International Conference on Music Retrieval* (pp. 89-94). Philadelphia: Drexel University.

Pearce, M., Müllensiefen, D. and Wiggins, G. (2010). Melodic grouping in music information retrieval: new methods and applications. In Z.W. Ras and A. Wieczorkowska (Eds.), *Advances in Music Information Retrieval* (pp. 364-388). Berlin: Springer.

Pinkerton, R. (1956). Information theory and melody. *Scientific American*, *194* (pp. 77-86).

Rissanen, J. (1978). Modeling by the shortest data description. *Automatica*, *14* (pp. 465-471).

Rissanen, J. (1987). Minimum-description-length principle. *Encyclopedia of Statistical Sciences*, *5* (pp. 523-527).

Rolland, P. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, *28*, No. 4 (pp. 334-350).

Schaffrath, H. (1995). The essen folksong collection. In D. Huron (Ed.), *Database containing 6,,255 folksong transcriptions in the Kern format and a 34-page research guide*. CCARH IC Menlo Park, CA.

Schenker, H. (1935/1979). In E. Oster (Ed.), *Free Composition*. New York: Longman.

Schenker, H. (1954). *Harmony*. London: University of Chicago Press.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27* (pp. 379-423 and 623-656).

Shannon, C. (1951). Prediction and entropy of printed english. *Bell Systems Technical Journal* (pp. 50-64).

Shannon, C. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.

Solomonoff, R. (1964a). A formal theory of inductive inference, part II. *Information and Control*, *7*, No. 2 (pp. 224-254).

Solomonoff, R. (1964b). A formal theory of inductive inference, part i. *Information and Control*, *7*, No. 1 (pp. 1-22).

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, Massachusetts: MIT Press.

Temperley, D. (2004). Bayesian models of musical structure and cognition. *Musicae Scientiae*, *8*, No. 2 (pp. 175-205).

Temperley, D. (2007). *Music and Probability*. Cambridge, Massachusetts: The MIT Press.

Temperley, D. (2008). A probabilistic model of melody perception. *32. Cognitive Science: A Multidisciplinary Journal*, No. 2 (pp. 418-444).

Tenney, J. and Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, *24* (pp. 205-241).

Thornton, C. (2010). Gauging the value of good data: informational embodiment quantification. *Adaptive Behavior*, *18*, No. 5 (pp. 389-399).

Toiviainen, P. and Eerola, T. (2002). A computational model of melodic similarity based on multiple representations and self-organizing maps. In C. Stevens, D. Burham, G. McPherson, E. Schubert and J. Rewick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition* (pp. 236-239). Sidney, Adelaide: Causal Productions.

Wertheimer, M. (1923/1938). Laws of orgnization in perceptual forms. In W. Ellis (Ed.), *A Source Book of Gestalt Psychology*. London: Routledge & Kegan Paul.

Youngblood, J. (1958). Style as information. *Journal of Music Theory*, *2* (pp. 24-35).